

TASK 1:

pwd: displays the current directory.

```
$ pwd
/Users/gift
$
```

vi : Starts the vi editor for editing a file

```
vi Testfile.txt
```

touch: touch file sets the current time stamp for a file. Can be used to create an empty file.

Examples below:

- Create a blank file

```
$ touch test.txt
ls
$ ls -l test.txt
-rw-r--r-- 1 gift staff 0 Mar 18 12:27 test.txt
```

- Change access time

```
$ touch -c test.txt
ls -l test.txt
-rw-r--r-- 1 gift staff 0 Mar 18 12:32 test.txt
```

- Create lots and lots of files

```
$ touch {A..Z}
$ touch {1..20}
$ touch {1..1000}.txt
```

```
$ ls
1      11     15     19     3      5      7      9      C      G      K      O      S      W
1.txt  12     16     2      3.txt  5.txt  7.txt  9.txt  D      H      L      P      T      X
10     13     17     2.txt  4      6      8      A      E      I      M      Q      U      Y
10.txt 14     18     20     4.txt  6.txt  8.txt  B      F      J      N      R      V      Z
```

mkdir : Create a new directory

```
$ mkdir gc_dir
$ ls -l
```

```
drwxr-xr-x 2 gift staff 64 Mar 18 12:46 gc_dir
```

rm : deletes a file or directory

```
$ ls -l
drwxr-xr-x 2 gift staff 64 Mar 18 12:46 gc_dir
-rw-r--r-- 1 gift staff 0 Mar 18 12:49 testfile.txt
$ rm testfile.txt
$ rm -r gc_dir/
$ ls -l
$
```

ls : displays the contents of a directory

```
$ ls -l
drwxr-xr-x 2 gift staff 64 Mar 18 12:46 gc_dir
-rw-r--r-- 1 gift staff 0 Mar 18 12:49 testfile.txt
```

echo : display a line of text/string on standard output or a file.

```
$ echo "display a line of text/string on standard output or a file."
display a line of text/string on standard output or a file.
```

cat : displays the content of a file on the standard output channel (usually the screen).

```
$ cat testfile.txt
This is a test file
re
fgtsss
ddd
$
```

who: display the users that are currently logged into your Unix computer system.

```
$ who
gift console Mar 17 21:57
gift ttys000 Mar 18 12:17
$
```

cd : Changes the working directory

```
$ pwd
```

```
/Users/gift
```

```
$cd /Users/gift/work
```

```
$pwd]
```

```
/Users/gift/work
```

date: displays the current system date and time

```
$ date
```

```
Mon Mar 18 14:47:54 CAT 2019
```

cal: he calendar of a specific month or a whole year.

```
$ cal
```

```
March 2019
```

```
Su Mo Tu We Th Fr Sa
```

```
1 2
```

```
3 4 5 6 7 8 9
```

```
10 11 12 13 14 15 16
```

```
17 18 19 20 21 22 23
```

```
24 25 26 27 28 29 30
```

```
31
```

```
$ cal 05 2006
```

May 2006

Su Mo Tu We Th Fr Sa

1 2 3 4 5 6

7 8 9 10 11 12 13

14 15 16 17 18 19 20

21 22 23 24 25 26 27

28 29 30 31

mv : moves file/directory to another folder

```
$ pwd
```

```
/Users/gift
```

```
ls
```

```
testfile.txt
```

```
ls work/
```

```
mv testfile.txt work/
```

```
ls work/
```

```
testfile.txt
```

cp: copies file/directory

```
ls
```

```
testfile.txt
```

```
cp testfile.txt testfile2.txt
```

```
ls
```

```
testfile.txt testfile2.txt
```

which: searches the path of executable in system paths set in \$PATH environment variable.

```
$ which cat
```

```
/bin/cat
```

```
$ which vi echo ls
```

```
/usr/bin/vi
```

/bin/echo

/bin/ls

TASK 2:

Word Mean

```
[cloudera@quickstart ~]$ hadoop fs -put testfile.txt /user/cloudera/wordcount/input/
[cloudera@quickstart ~]$ hadoop fs -ls /user/cloudera/wordcount/input
Found 1 items
-rw-r--r-- 1 cloudera cloudera 30540 2019-03-18 08:16 /user/cloudera/wordcount/input/testfile.txt
[cloudera@quickstart ~]$
```

```
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ hadoop jar /usr/lib/hadoop-mapreduce/hadoop-mapreduce-examples-2.6.0-cdh5.13.0.jar wordcount wordcount/input
wordcount/output
19/03/18 08:40:18 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
19/03/18 08:40:19 INFO input.FileInputFormat: Total input paths to process : 1
19/03/18 08:40:19 INFO mapreduce.JobSubmitter: number of splits:1
19/03/18 08:40:19 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1552919899841_0002
19/03/18 08:40:20 INFO impl.YarnClientImpl: Submitted application application_1552919899841_0002
19/03/18 08:40:20 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1552919899841_0002/
19/03/18 08:40:20 INFO mapreduce.Job: Running job: job_1552919899841_0002
19/03/18 08:40:31 INFO mapreduce.Job: Job job_1552919899841_0002 running in uber mode : false
19/03/18 08:40:31 INFO mapreduce.Job: map 0% reduce 0%
19/03/18 08:40:39 INFO mapreduce.Job: map 100% reduce 0%
19/03/18 08:40:45 INFO mapreduce.Job: map 100% reduce 100%
19/03/18 08:40:45 INFO mapreduce.Job: Job job_1552919899841_0002 completed successfully
19/03/18 08:40:45 INFO mapreduce.Job: Counters: 49
File System Counters
  FILE: Number of bytes read=19739
  FILE: Number of bytes written=326747
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=30679
  HDFS: Number of bytes written=16669
  HDFS: Number of read operations=6
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=4071
  Total time spent by all reduces in occupied slots (ms)=3404
  Total time spent by all map tasks (ms)=4071
  Total time spent by all reduce tasks (ms)=3404
  Total vcore-milliseconds taken by all map tasks=4071
  Total vcore-milliseconds taken by all reduce tasks=3404
  Total megabyte-milliseconds taken by all map tasks=4168704
  Total megabyte-milliseconds taken by all reduce tasks=3485696
Map-Reduce Framework
  Map input records=223
  Map output records=2278
  Map output bytes=36008
  Map output materialized bytes=19739
  Input split bytes=139
  Combine input records=2278
  Combine output records=765
  Reduce input groups=765
  Reduce shuffle bytes=19739
  Reduce input records=765
  Reduce output records=765
  Spilled Records=1530
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=69
  CPU time spent (ms)=1740
  Physical memory (bytes) snapshot=467881984
  Virtual memory (bytes) snapshot=3153989632
  Total committed heap usage (bytes)=390070272
Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0
File Input Format Counters
```

```

        Bytes Read=30540
        File Output Format Counters
        Bytes Written=16669
[cloudera@quickstart ~]$

[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ hadoop fs -ls /user/cloudera/wordcount/output
Found 2 items
-rw-r--r-- 1 cloudera cloudera 0 2019-03-18 08:40 /user/cloudera/wordcount/output/_SUCCESS
-rw-r--r-- 1 cloudera cloudera 16669 2019-03-18 08:40 /user/cloudera/wordcount/output/part-r-00000
[cloudera@quickstart ~]$

```

Word Mean

```

[cloudera@quickstart ~]$ hadoop fs -put testfile.txt /user/cloudera/wordmean/input
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ hadoop jar /usr/lib/hadoop-mapreduce/hadoop-mapreduce-examples-2.6.0-cdh5.13.0.jar wordmean wordmean/input
wordmean/output
19/03/18 09:01:36 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032
19/03/18 09:01:36 WARN security.UserGroupInformation: PrivilegedActionException as:cloudera (auth:SIMPLE)
cause:org.apache.hadoop.mapred.FileAlreadyExistsException: Output directory
hdfs://quickstart.cloudera:8020/user/cloudera/wordmean/output already exists
org.apache.hadoop.mapred.FileAlreadyExistsException: Output directory hdfs://quickstart.cloudera:8020/user/cloudera/wordmean/output
already exists
    at org.apache.hadoop.mapreduce.lib.output.FileOutputFormat.checkOutputSpecs(FileOutputFormat.java:146)
    at org.apache.hadoop.mapreduce.JobSubmitter.checkSpecs(JobSubmitter.java:270)
    at org.apache.hadoop.mapreduce.JobSubmitter.submitJobInternal(JobSubmitter.java:143)
    at org.apache.hadoop.mapreduce.Job$10.run(Job.java:1307)
    at org.apache.hadoop.mapreduce.Job$10.run(Job.java:1304)
    at java.security.AccessController.doPrivileged(Native Method)
    at javax.security.auth.Subject.doAs(Subject.java:422)
    at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1917)
    at org.apache.hadoop.mapreduce.Job.submit(Job.java:1304)
    at org.apache.hadoop.mapreduce.Job.waitForCompletion(Job.java:1325)
    at org.apache.hadoop.examples.WordMean.run(WordMean.java:186)
    at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:70)
    at org.apache.hadoop.examples.WordMean.main(WordMean.java:163)
    at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
    at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
    at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
    at java.lang.reflect.Method.invoke(Method.java:498)
    at org.apache.hadoop.util.ProgramDriver$ProgramDescription.invoke(ProgramDriver.java:71)
    at org.apache.hadoop.util.ProgramDriver.run(ProgramDriver.java:144)
    at org.apache.hadoop.examples.ExampleDriver.main(ExampleDriver.java:74)
    at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
    at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
    at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
    at java.lang.reflect.Method.invoke(Method.java:498)
    at org.apache.hadoop.util.RunJar.run(RunJar.java:221)
    at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
[cloudera@quickstart ~]$ hadoop fs -rm -r /user/cloudera/wordmean/output
Deleted /user/cloudera/wordmean/output
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ hadoop jar /usr/lib/hadoop-mapreduce/hadoop-mapreduce-examples-2.6.0-cdh5.13.0.jar wordmean wordmean/input
wordmean/output
19/03/18 09:02:21 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032
19/03/18 09:02:22 INFO input.FileInputFormat: Total input paths to process : 1
19/03/18 09:02:22 INFO mapreduce.JobSubmitter: number of splits:1
19/03/18 09:02:22 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1552919899841_0005
19/03/18 09:02:22 INFO impl.YarnClientImpl: Submitted application application_1552919899841_0005
19/03/18 09:02:22 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1552919899841_0005/
19/03/18 09:02:22 INFO mapreduce.Job: Running job: job_1552919899841_0005
19/03/18 09:02:29 INFO mapreduce.Job: Job job_1552919899841_0005 running in uber mode : false
19/03/18 09:02:29 INFO mapreduce.Job: map 0% reduce 0%
19/03/18 09:02:35 INFO mapreduce.Job: map 100% reduce 0%
19/03/18 09:02:41 INFO mapreduce.Job: map 100% reduce 100%
19/03/18 09:02:42 INFO mapreduce.Job: Job job_1552919899841_0005 completed successfully
19/03/18 09:02:42 INFO mapreduce.Job: Counters: 49
    File System Counters
        FILE: Number of bytes read=39
        FILE: Number of bytes written=287343
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=30678
        HDFS: Number of bytes written=24
        HDFS: Number of read operations=6
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=2
    Job Counters
        Launched map tasks=1
        Launched reduce tasks=1
        Data-local map tasks=1
        Total time spent by all maps in occupied slots (ms)=4179
        Total time spent by all reduces in occupied slots (ms)=3245
        Total time spent by all map tasks (ms)=4179
        Total time spent by all reduce tasks (ms)=3245
        Total vcore-milliseconds taken by all map tasks=4179
        Total vcore-milliseconds taken by all reduce tasks=3245
        Total megabyte-milliseconds taken by all map tasks=4279296
        Total megabyte-milliseconds taken by all reduce tasks=3322880
    Map-Reduce Framework
        Map input records=223
        Map output records=4556
        Map output bytes=66062
        Map output materialized bytes=39

```

```

        Input split bytes=138
        Combine input records=4556
        Combine output records=2
        Reduce input groups=2
        Reduce shuffle bytes=39
        Reduce input records=2
        Reduce output records=2
        Spilled Records=4
        Shuffled Maps =1
        Failed Shuffles=0
        Merged Map outputs=1
        GC time elapsed (ms)=50
        CPU time spent (ms)=1720
        Physical memory (bytes) snapshot=469241856
        Virtual memory (bytes) snapshot=3145502720
        Total committed heap usage (bytes)=389545984

Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0

File Input Format Counters
    Bytes Read=30540
File Output Format Counters
    Bytes Written=24
The mean is: 10.802019315188762
[cloudera@quickstart ~]$ hadoop fs -ls /user/cloudera/wordmean/output
Found 2 items
-rw-r--r-- 1 cloudera cloudera 0 2019-03-18 09:02 /user/cloudera/wordmean/output/ SUCCESS
-rw-r--r-- 1 cloudera cloudera 24 2019-03-18 09:02 /user/cloudera/wordmean/output/part-r-00000
[cloudera@quickstart ~]$

```

WordStandardDeviation

```

[cloudera@quickstart ~]$ hadoop fs -put testfile.txt /user/cloudera/wordstandarddeviation/input
[cloudera@quickstart ~]$ hadoop jar /usr/lib/hadoop-mapreduce/hadoop-mapreduce-examples-2.6.0-cdh5.13.0.jar wordstandarddeviation
wordstandarddeviation/input wordstandarddeviation/output
19/03/18 09:21:39 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
19/03/18 09:21:40 INFO input.FileInputFormat: Total input paths to process : 1
19/03/18 09:21:40 INFO mapreduce.JobSubmitter: number of splits:1
19/03/18 09:21:40 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1552919899841_0007
19/03/18 09:21:40 INFO impl.YarnClientImpl: Submitted application application_1552919899841_0007
19/03/18 09:21:40 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1552919899841_0007/
19/03/18 09:21:40 INFO mapreduce.Job: Running job: job_1552919899841_0007
19/03/18 09:21:47 INFO mapreduce.Job: Job job_1552919899841_0007 running in uber mode : false
19/03/18 09:21:47 INFO mapreduce.Job: map 0% reduce 0%
19/03/18 09:21:53 INFO mapreduce.Job: map 100% reduce 0%
19/03/18 09:21:59 INFO mapreduce.Job: map 100% reduce 100%
19/03/18 09:22:00 INFO mapreduce.Job: Job job_1552919899841_0007 completed successfully
19/03/18 09:22:00 INFO mapreduce.Job: Counters: 49

File System Counters
    FILE: Number of bytes read=56
    FILE: Number of bytes written=287589
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=30691
    HDFS: Number of bytes written=39
    HDFS: Number of read operations=6
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2

Job Counters
    Launched map tasks=1
    Launched reduce tasks=1
    Data-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=3601
    Total time spent by all reduces in occupied slots (ms)=3474
    Total time spent by all map tasks (ms)=3601
    Total time spent by all reduce tasks (ms)=3474
    Total vcore-milliseconds taken by all map tasks=3601
    Total vcore-milliseconds taken by all reduce tasks=3474
    Total megabyte-milliseconds taken by all map tasks=3687424
    Total megabyte-milliseconds taken by all reduce tasks=3557376

Map-Reduce Framework
    Map input records=223
    Map output records=6834
    Map output bytes=100232
    Map output materialized bytes=56
    Input split bytes=151
    Combine input records=6834
    Combine output records=3
    Reduce input groups=3
    Reduce shuffle bytes=56
    Reduce input records=3
    Reduce output records=3
    Spilled Records=6
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=473
    CPU time spent (ms)=2760
    Physical memory (bytes) snapshot=642076672
    Virtual memory (bytes) snapshot=3164168192
    Total committed heap usage (bytes)=546832384

Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0

```

```
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=30540
File Output Format Counters
  Bytes Written=39
The standard deviation is: 27.04312972098201
[cloudera@quickstart ~]$
```